

> Develop Predictive Models Using Your Business Expertise

Clementine 8.0 is a widely used, integrated workbench that facilitates the entire data mining process. It enables data miners and analysts to quickly access and prepare data for modeling, build, and rapidly compare models, and immediately deploy models into a real-time operational environment. Clementine provides decision makers with the ability to seamlessly integrate it with other business processes and systems to gain a complete view of current conditions and shed light on future events.

Confidently and productively explore your data

Clementine's workflow interface enables you to add your business knowledge—the key to successful data mining—to your process. Its highly visual user interface makes “train of thought” analysis possible. Clementine produces streams, the visual map of your data mining steps, to automatically document the analytical process. You can quickly explore hunches or ideas by interacting with the stream. This makes it possible for you to focus on knowledge discovery rather than on performing technical tasks, such as writing code, and reach results in the shortest amount of time possible.

Enhancements to Clementine 8.0 further improve your data mining productivity. You can visually assess and profile large, complex datasets using the new data audit node, and effectively reduce the number of required data exploration steps. New data manipulation nodes include binning and reclassification capabilities for faster data preparation. The new cluster viewer, which provides a visual summary of discovered clusters, enables you to more thoroughly explore and understand clusters.

Select from an unparalleled breadth of techniques

Clementine offers a full range of machine-learning and statistical techniques for clustering, classification, association, and prediction. SPSS Inc. has more than 35 years experience in predictive analytics, and its algorithms are verified and calibrated to deliver powerful data mining results. You can be assured of having the depth and breadth of analytical techniques needed to attain the best results for all of your data mining tasks.

Put results to work in your organization

SPSS offers several deployment options to meet your needs for real-time or in-database scoring. Clementine exports all data mining steps, including data access, modeling, and post-processing—not just the model—to save your organization time and money. Save predictive model markup language (PMML), an industry standard for exchanging predictive models, and embed it into your system. You can publish processes or streams and transparently deliver results to other software applications.

Easily integrate Clementine into your existing systems

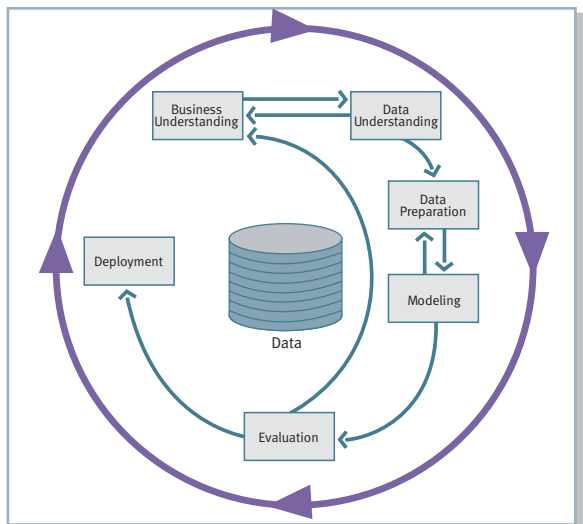
With Clementine, there's good news for your IT staff. Because it's an open, standards-based system, Clementine easily integrates into your organization's existing operational systems and databases. Clementine leverages your organization's database investment by pushing much of the data preprocessing tasks, typically 80-90 percent of your data mining efforts, back into the database. And because preprocessing occurs on the database tier and not the server tier or the client tier, network traffic is minimized.

Specifications*

CRISP-DM

- Help your company focus data mining on problem solving rather than the technology. Clementine supports CROss-Industry Standard Process for Data Mining (CRISP-DM), the de facto industry standard, at every step of the process. To support CRISP-DM, Clementine includes a project manager that enables you to easily focus on the particular steps of your data mining projects. New capabilities, such as the drag-and-drop tool, make the project manager even easier to use in Clementine 8.0.

The diagram below shows the CRISP-DM process. Following the diagram, Clementine's capabilities for each step are summarized, along with a brief discussion of other features that help make Clementine the data mining workbench of choice.



The CRISP-DM phases, as shown in the diagram above, enable you to focus on implementing data mining projects that yield real business results.

Business understanding

- Apply proven methodology and best practices to your data mining projects using business-specific expertise available through the optional Clementine Application Templates (CATs). Use these templates (listed below) with Clementine to accelerate your data mining projects.
 - CRM CAT
 - Web mining CAT
 - Telco CAT
 - Fraud CAT
 - Microarray CAT

Data understanding

- Generate subsets of data automatically from graphs and tables
- Show summary statistics, histograms, and distribution graphics for each data field, and display them in an easy-to-read matrix with the data audit node. This provides you with a comprehensive first look at your data.
- Visually interact with your data
 - Select node or field and view information in a table
 - Create histograms, distributions, line plots, and point plots
 - Display 3-D, panel, and animated graphs
 - Use Web association detection

Data preparation

- Bring data into Clementine using a variety of methods and sources
 - Data access
 - Use the security of your selected database or operating system to ensure data security
 - Access any ODBC-compliant data source. You can use the included SPSS Data Access Pack to work with data from Oracle®, Microsoft® SQL Server, and IBM® DB2®
 - Import delimited and fixed-width text, SPSS®, and SAS® 6, 7, 8, and 9 files
 - Data output
 - Work with delimited and fixed-width text, ODBC, SPSS, Microsoft® Excel, and SAS 6, 7, 8, and 9 files
 - Text access
 - Use Text Mining for Clementine™ to automatically extract concepts from database notes and documents
 - Web mining
 - Use Web Mining for Clementine™ to automatically extract Web site events from Web logs.**

- Choose from various data-cleaning options
 - Remove or replace invalid characters through the variable file and fixed file source nodes
 - Automatically fill in missing values
- Manipulate data
 - Work with complete record and field operations, including:
 - Field filtering, naming, derivation, and value replacement
 - Record selection, sampling, merging and concatenation, sorting, aggregation and balancing, deriving new fields based on conditional criteria, and calculating new fields

- Specialized manipulations for showing the “history” of values and converting set variables into flag variables
- Automatically create new set fields based on the values of one or more existing numeric range fields with the binning node
- Re-group or collapse categories for numeric or string set fields with the reclassify node. This node enables the transformation from one set of discrete values to another.
- Customize the natural order of fields downstream with the field reorder node
- Merge records using the following types of joins:
 - Inner join
 - Full outer join
 - Partial outer join, both left and right joins
 - Anti-join (the opposite of an inner join)

Modeling

- Prediction and classification
 - Neural networks (multi-layer perceptrons trained using error-back propagation with momentum, radial basis function, and Kohonen network)
 - Decision trees and rule induction [C5.0 and Classification and Regression Trees (C&RT)]
 - Linear regression, logistic regression, and multinomial logistic regression
- Clustering and segmentation
 - Kohonen network, K-means, and TwoStep
 - View summary statistics and distributions for fields between clusters using the Cluster Viewer
- Association detection
 - GRI, apriori, and sequence
- Data reduction
 - Factor analysis and principle components analysis
- Models can be quickly combined for greater accuracy
 - Meta-modeling easily enables predictions from one model to be used directly for building a second model
- Clementine External Module Interface (CEMI) for custom algorithms
 - Add data access, modeling, and graphics
 - Create and include node palettes and generated model palettes using custom CEMI icons
 - Import PMML-generated models created in other tools

- Algorithm User Manual, which explains the theories and methods of the algorithms in Clementine

Evaluation

- Easily evaluate models using lift, gains, profit, and response graphs
 - Use a one-step process that shortens project time when evaluating multiple models
- Define hit conditions and scoring expressions to interpret model performance

Deployment

- Clementine Solution Publisher (optional)
 - Automate the export of all operations, including data access; data manipulations; model scoring, including combinations of models; and post-processing
 - Work in a runtime environment for executing image files on target platforms
 - Easily update solutions through small image file
- Cleo™ (optional)
 - Rapidly deploy predictive models online through an XML-based tool
 - Enable multiple users to simultaneously access and immediately score single records, multiple records, or an entire database using the customizable browser-based interface
- Clementine Batch Mode
 - Automate production tasks, working outside the user interface. With Clementine Batch Mode, you can:
 - Automate Clementine processes from other applications or scheduling systems. Just implement Clementine in batch mode to execute a script.
 - Generate encoded passwords
 - Launch commands using a command file
 - Execute selected lines from a stream, SuperNode, or stand-alone script using an icon on the toolbar
 - Update stream parameters within a standalone script
 - Refresh source nodes automatically upon stream execution
- Scripting
 - Automate command line scripts or scripts associated with Clementine streams to automate repetitive tasks in the user interface. Scripts generally perform the same type of actions that you can do with a mouse or keyboard.

- Export your generated models as PMML
- Perform in-database scoring
 - Reduce costly data transfer and calculations on the client when you score data in the database
 - Generate SQL from linear regression and factor-generated models
 - Generate SQL from decision trees and rule sets
- Use the bulk loader program
 - Increase performance during data export by using a bulk loader program that's specific to your database. Fine-tune various options, including row-size or column-wise binding for loading via ODBC, and batch-size setting for batch commits to the database.

Scalability and integration

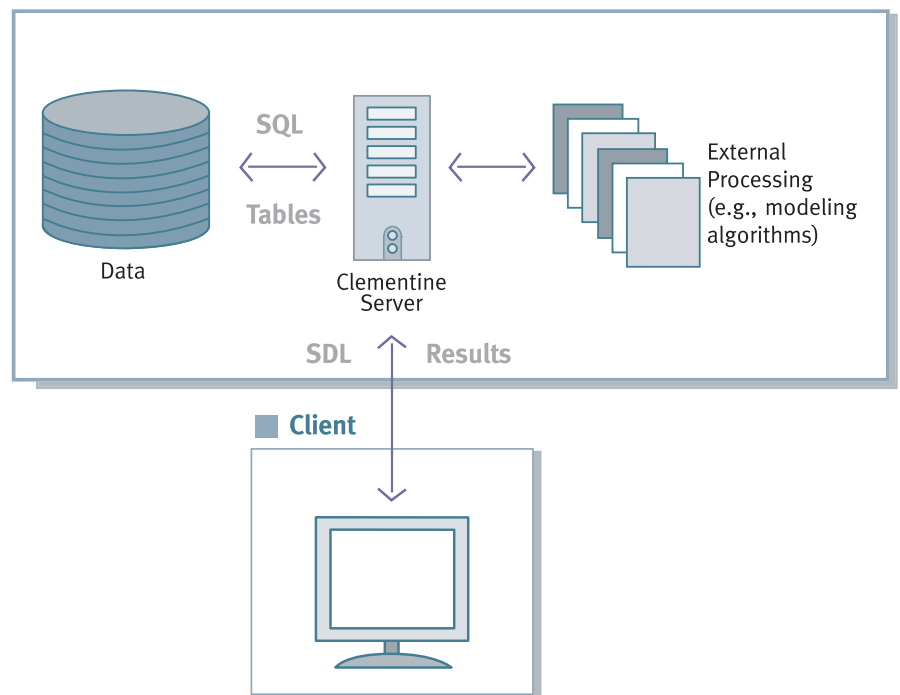
Scalability

- Use in-database mining and server-tier processing to leverage parallel database implementations
- Minimize network traffic via intelligent field projection—data are pulled as needed from your data warehouse and pass only relevant results to the client

Integration

- Clementine 8.0 features enhanced integration with other SPSS Inc. products, including:
 - SPSS for Windows®, which enables you to launch SPSS Server directly from Clementine Server
 - AnswerTree®
 - SPSS Scenario Manager Suite technology
 - Wizard support for Cleo
- Clementine 8.0 also provides integration with databases, market automation tools, and other software tools through CEMI

Server



Clementine uses a highly scalable, three-tiered architecture for data mining, as shown in this diagram. The Clementine Client tier (shown at the bottom of the illustration) passes stream description language (SDL) to Clementine Server. Clementine Server then analyzes these particular tasks to determine which tasks it can push to the database. After the database runs the tasks it can process, it passes only the relevant aggregated tables to Clementine Server. If you are using a CEMI, such as Text Mining for Clementine or Capri™, Clementine Server passes the relevant tasks to that particular external process.

System requirements

Clementine Client:

- Operating system: Microsoft® Windows® Me, Windows XP Home or Professional, Windows 2000, Windows 2003, or Windows NT® 4.0 with Service Pack 6
- Minimum RAM: 256MB are recommended. 512MB are recommended when using CATs or other large datasets.
- Minimum free disk space: 320MB
- A CD-ROM drive is required for installation
- Web browser: Internet Explorer version 5.x or Netscape® 6, for running the online help system

Clementine Server, Clementine Solution

Publisher Runtime (optional), and Clementine Batch Mode:

- Operating system: Windows 2000 Professional or Advanced Server; Windows NT 4.0 with Service Pack 6 or later; SunSM Solaris™ 2.7, 2.8, or 2.9; HP-UX 11i; IBM® AIX® 4.3.3 or AIX 5L version 5.1 or later; or IBM OS/400™ (iSeries™) V5R2 with OS/400 Portable Applications Solution Environment (PASE, 5722-SS1 Option 33)
- Hardware: Pentium®-compatible processor, SPARC® for Solaris (Sun UltraSPARC® 2 hardware or better if running on Solaris), HP Workstation for HP-UX (PA-RISC processor), HP 9000 server if running on HP-UX, IBM RS/6000® for AIX (PowerPC® 233MHz or better if running on AIX), or IBM eServer iSeries™ for OS/400
- Minimum RAM: 256MB; 512MB recommended
- Minimum free drive space: 128MB of available disk space are required for installation. Additional free disk space is required to run the program (for temporary files). 1GB is recommended. For Clementine Solution Publisher Runtime, the minimum free disk space required to install the software is 64MB, plus at least twice the disk space of the amount of data to be processed.
- A network adapter running TCP/IP protocol
- A CD-ROM drive is required for installation

Cleo (optional):

- Web server: Cleo requires at least one server computer that meets the following minimum requirements. Using additional processors, faster processors, and more RAM will improve performance.
 - Operating system: Windows 2000 Server, Windows NT 4.0 Server with Service Pack 5 (cannot be installed on Windows NT Terminal Server), or Sun Solaris Version 7 or later with UltraSPARC 2 hardware or better
 - Hardware: 500MHz processor or faster
 - Minimum RAM: 512MB
 - Minimum free drive space: 700MB of available disk space
 - A graphics adapter with 800 x 600 resolution (SVGA) or higher, capable of displaying at least 256 colors
 - A network adapter running the TCP/IP protocol
- Repository: the system requires a database to serve as a repository for published content, framework settings, and other information. The following databases are supported:
 - Microsoft SQL Server 2000
 - Oracle 8i, version 8.1.7
- Data warehouse: the system can be configured to access data from a data warehouse or database. The system has only been tested with SQL Server 2000 and Oracle 8i databases.
- Web client: content is delivered to clients as standard HTML pages. Supported browsers include:
 - Internet Explorer version 5.5 with Service Pack 2 or version 6.0 for Windows
 - Internet Explorer version 5.2 for Macintosh®
 - Netscape 6.2

Text Mining for Clementine (optional):

- Clementine 7.1 or above
- Operating system: Windows NT Server 4.0 with Service Pack 6 or later or Windows 2000 Server or Professional
- Minimum RAM: 256MB
- Minimum free disk space: 85MB, plus space for databases
- Web Browser: Internet Explorer 5.0 or later or Netscape 6.0 or later is required to use the Viewer node

Web Mining for Clementine (optional):

- Clementine 7.2 or later
- Operating system: Windows XP or Windows 2000 Server or Professional
- Minimum RAM: 512MB
- Minimum free disk space: two times the amount of raw Web data being processed

Clementine Application Templates (optional):

- Clementine 7.2
- Minimum RAM: 512MB are recommended

Capri for Clementine (optional):

Plug-in algorithm for detecting sequences

- Operating system: Windows 98, 2000, or Windows NT 4.0 with Service Pack 3 or later, or Sun Solaris 2.6
- Hardware: Pentium-compatible processor or higher for Windows or SPARC for Solaris
- Minimum RAM: 128MB
- Minimum free drive space: 3MB
- A CD-ROM drive is required for installation

To learn more, please visit www.spss.com. For SPSS office locations and telephone numbers, go to www.spss.com/worldwide.

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners. © Copyright 2003 Integrated Solutions Ltd. CLM8SPC-1103

The SPSS logo, consisting of the letters "SPSS" in a bold, white, sans-serif font, set against a red square background.